**PROJECT PROPOSAL**

Vinay Narendra Gurrap – 13475101

Email ID – [gurrapv@coventry.ac.uk](mailto:gurrapv@coventry.ac.uk)

* **Project Title** -Customer Churn Prediction for a Telecommunication Company
* **Project Problem**

The aim of this project is to build a predictive model which will help in identifying potential churners within a telecom company’s customer base. The issue of the customer churn is a major concern for telecom companies as it has a significant impact on their revenue and profits. Customers who are more likely to churn needs to be predicted and identified early on by taking proactive steps to retain them with better customer services, tailored offers and targeted promotions.

* **Planned Tasks**
  1. *Business Understanding* – The telecom industry faces fierce competition in satisfying its customers and customer churn is one of the biggest fears of the industry. Churn rate is the number of customers who quit their service or the number of employees who quit their jobs. Understanding the business implications of customer churn is important and minimizing it is essential to achieve growth and success in this competitive industry.
  2. *Data Understanding –* It is important to have a high-quality dataset for data understanding as the models will be built on it. The dataset will be examined thoroughly to make sure that the dataset quality is good.
  3. *Data Preparation* – This phase involves identifying desired attributes or variables., cleaning the data by handling null values, deleting duplicate variables, handling outliers, changing the data type to their required type which will help in preparing the data for modelling.
  4. *Modelling* – Develop various classification machine learning models such as KNN, Random Forest, Decision Tree, etc. in addition to implementing ensemble models to find out the best model for predicting behavior to retain customers by dividing the dataset into train, test and validation sets and adjusting hyperparameters to enhance the model’s performance.
  5. *Evaluation and Deployment* – Use relevance measures such as Accuracy, F1-Score, Precision, Recall and AUC-ROC curve to assess the effectiveness of the machine learning models and selecting the model effective model based on the results.
* **Dataset**

The dataset that is being used for this project is “Telecom Customer Churn Prediction”. It is available of Kaggle and can be accessed using the following link: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

* **Work Plan**

Weeks 1 – 2: *Related Work Analysis, Data Acquisition, Data Pre-Processing and Exploratory Data Analysis*:

* + Refer existing papers, theories and machine learning techniques for customer churn prediction.
  + Identifying the relevant dataset to be acquired, perform the initial data exploration to understand the dataset, pre-processing and cleaning the dataset to find missing values, perform feature selection if needed and EDA to gain insights and understand customer behavior and patterns leading to potential churn factors by using PySpark for data analysis and Tableau for data visualization.

Weeks 3 – 4: *Feature Engineering, Model Selection, Optimization and Evaluation and Research Report Documentation:*

* + Improve the accuracy of the predicting model by selecting relevant features which gives meaningful insights.
  + Implement various classification machine learning algorithms to predict customer churn and validate the ensemble models using appropriate validation techniques.
  + Evaluate the model performance through classification report and confusion matrix and optimize it through parameter tuning.
  + Compare the performance of all the models, derive insights into the factors that contributes to the customer churn.
  + Prepare the research report with the findings and insights, finalizing the visualizations.
  + Write the conclusion of the research report, mention future suggestions to improve the research project and social impact of the project.